Supplementary Material

related to 'A Systematic Review of Large Language Models in Medical Specialties: Applications, Challenges and Future Directions.'

Asma Musabah Alkalbani, Ahmed Salim Alrawahi, Ahmad Salah, Venus Haghighi,
Yang Zhang, Salam Alkindi, Quan Z Sheng

Table 1: Quality Assessment Criteria

Domain	**Subject selection**	**Index text (LLM)**	**Reference standard**	**Work-flow**
Signaling questions	Q1: Were the input data well described (source, size, and quality) Q2: Were the subject selection criteria clearly presented? Q3: Was data pre-processing performed?	Q4: Were the split and justification of training and test datasets provided? Q5: Was external verification performed?	Q6: Was the reference standard likely to correctly perform the intended task?	Q7: Was the LLM's performance evaluated in one of the following 1) real-time 2) near-real-time medical settings, 3) pre-existing clinical cases?
Description of risk of bias concerns	The quality assessment evaluates the risk of bias as low, high, or unclear. In case all signaling questions were answered yes, the risk of bias is low. In case any signaling question was answered no, the potential for bias is highlighted for further discussions by reviewers to evaluate the risk of bias. In case there is insufficient data reported by a reviewed study, the risk of bias is judged as unclear.			

Table 2: QUADAS-AI Quality assessment results

Authors, Year	Subject Selection				Index Test (LLM)			Reference Standard		Flow and Timing	
	Q1	Q2	Q3	Risk	Q4	Q5	Risk	Q6	Risk	Q7	Risk
Acharya, et al, 2024.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Ali, et al, 2023.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Alkouz, et al, 2022.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Ayoub, et al, 2024.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Bellinger, et al, 2023.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Bombieri, et al, 2023.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Bressem, et al, 2024.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Bushuven, et al, 2023.	Yes	Yes	Yes	Low	No	No	High	Yes	Low	Yes	Low
Butler, et al, 2024.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Cai, et al, 2021.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Campbell, et al, 2024.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Caruccio, et al, 2024.	Yes	No	No	High	Yes	Yes	Low	Yes	Low	Yes	Low
CB, et al, 2023.	Yes	No	Yes	Low	No	Yes	High	Yes	Low	Yes	Low
Chen, et al, 2021.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Chizhikova, et al, 2023.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Cuthbert, et al, 2023.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Datta, et al, 2022.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Datta, et al, 2023.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
De Oliveira, et al, 2024.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Fonseca, et al, 2024.	Yes	No	Yes	High	No	Yes	Low	Yes	Low	Yes	Low
Fu, et al, 2022.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Gan, et al, 2024.	Yes	Yes	No	Low	No	Yes	Low	Yes	Low	Yes	Low
Ghanem, et al, 2024.	No	No	No	High	No	No	High	Yes	Low	Yes	Low
Guo, et al, 2022.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Hartman, et al, 2023.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Haze, et al, 2023.	Yes	No	No	Low	No	No	High	Yes	Low	Yes	Low
Homburg, et al, 2023.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Hristidis, et al, 2023.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low

Table 2: QUADAS-AI Quality assessment results

Authors, Year	Subject Selection				Index Test (LLM)			Reference Standard		Flow and Timing	
	Q1	Q2	Q3	Risk	Q4	Q5	Risk	Q6	Risk	Q7	Risk
Hu, et al, 2024.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Ilias, et al, 2023.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Ji, et al, 2021.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Kaarre, et al, 2023.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Kaplar, et al, 2022.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Karakas, et al, 2023.	No	No	No	High	No	No	High	No	High	Yes	Low
Karthikeyan, et al, 2022.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Kim, et al, 2023. under	Yes	No	Yes	Low	No	Yes	High	Yes	Low	Yes	Low
Kim, et al, 2023. predi	Yes	No	Yes	Low	Yes	Yes	High	Yes	Low	Yes	Low
Klein, et al, 2021.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Laison, et al, 2023.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Lee, et al, 2024.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Li, et al, 2021.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Li, et al, 2024.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Liu, et al, 2021.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Liu, et al, 2023. using	Yes	Yes	No	Low	No	Yes	Low	Yes	Low	Yes	Low
Liu, et al, 2023.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Liu, et al, 2024.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Lu, et al, 2021. natu	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Lu, et al, 2021. Revea	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Maillard, et al, 2023.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Massey, et al, 2023.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Mihalache, et al, 2023.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Mika, et al, 2023.	Yes	No	No	Low	No	Yes	Low	Yes	Low	Yes	Low
Mishra, et al, 2023.	Yes	No	No	Low	No	Yes	Low	Yes	Low	Yes	Low
Mitchell, et al, 2022.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Mohamad, et al, 2023.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Mulyar, et al, 2021.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low

Table 2: QUADAS-AI Quality assessment results

Authors, Year	Subject Selection				Index Test (LLM)			Reference Standard		Flow and Timing	
	Q1	Q2	Q3	Risk	Q4	Q5	Risk	Q6	Risk	Q7	Risk
Nimmi, et al, 2022.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Oon, et al, 2023.	Yes	No	No	Low	No	Yes	Low	Yes	Low	Yes	Low
Otsuka, et al, 2023.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Parker, et al, 2023.	No	No	No	High	No	No	High	No	High	Yes	Low
Percha, et al, 2021.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Qiao, et al, 2021.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Scquizzato, et al, 2024.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Shang, et al, 2023.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Song, et al, 2023.	Yes	No	No	Low	No	Yes	Low	Yes	Low	Yes	Low
Tan, et al, 2023.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Truhn, et al, 2023.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Tsoutsanis, et al, 2024.	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Vithanage, et al, 2024.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Walker, et al, 2023.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Wang, et al, 2022.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Wang, et al, 2023. perfo	Yes	No	Yes	Low	No	Yes	Low	No	High	Yes	Low
Wang, et al, 2023. poten	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Wang, et al, 2023. chtgpt	Yes	No	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Wang, et al, 2023. text	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Wang, et al, 2023. effici	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Not clear	Not clear
Wilhelm, et al, 2023.	Yes	No	No	Low	No	Yes	Low	Yes	Low	Yes	Low
Xie, et al, 2022.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Xue, et al, 2024.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Yun, et al, 2023.	Yes	No	No	Low	No	Yes	Low	Yes	Low	Yes	Low
Zaboli, et al, 2024.	Yes	Yes	Yes	Low	No	Yes	Low	Yes	Low	Yes	Low
Zhang, et al, 2022.	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Zhang, et al, 2022.Lab	Yes	No	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low
Zhou, et al, 2022.	Yes	Yes	Yes	Low	Yes	Yes	Low	Yes	Low	Yes	Low

1 Search Strategy

Web of Science Query:
TS=((”large language models” OR LLMs OR GPT-2 OR GPT-3 OR GPT-3.5 OR GPT-4 OR ChatGPT OR BERT OR LAMDA OR RoBERTa OR ”Turing-NLG” OR ”Turing Bletchley” OR XLNet OR ”Transformer-XL” OR ERNIE OR ELECTRA OR ALBERT) AND (clinical OR disease OR healthcare OR medical OR patient OR diagnosis OR therapy OR treatment OR surgical OR oncology OR cardiology OR neurology OR psychological OR immunology OR genomics OR bioinformatics OR ”public health”))

Scopus Query:
((large AND language AND models) OR llms OR gpt-2 OR gpt-3 OR gpt-3.5 OR gpt-4 OR chatgpt OR bert OR lamda OR roberta OR turing-nlg OR (turing AND bletchley) OR xlnet OR transformer-xl OR ernie OR electra OR albert) AND (clinical OR disease OR healthcare OR medical OR patient OR diagnosis OR therapy OR treatment OR surgical OR oncology OR cardiology OR neurology OR psychological OR immunology OR genomics OR bioinformatics OR (public AND health))